

関連記事を用いた市況分析コメントの自動生成

Automatic generation of market analysis comments by using relevant articles

酒井 浩之^{1*} 坂地 泰紀² 和泉 潔² 松井 藤五郎³ 入江 圭太郎^{4†}
Hiroyuki Sakai¹ Hiroki Sakaji² Kiyoshi Izumi² Tohgoroh Matsui³ Keitaro Irie⁴

¹ 成蹊大学¹ Seikei University ² 東京大学² The University of Tokyo

³ 中部大学³ Chubu University

⁴ 三菱UFJ国際投信⁴ Mitsubishi UFJ Kokusai Asset Management

Abstract: 本研究では、経済新聞記事などの経済テキストから、日経平均株価などの市況について言及している記事を抽出し、それらの内容を自動的に要約することによりマーケットレポートにおける市況分析コメントを自動生成する手法の開発を行う。しかし、日経平均株価の市況について言及している記事のみでは、指定した期間において重要な内容について言及している文の数が少なく、そのために重要な内容が市況分析コメントに含まれないことがある。そこで本研究では、ある期間において日経平均株価に影響を与えたイベントを推定し、そのイベントについて述べた記事を関連記事として自動的に検索し、検索された関連記事をも使用することで、生成する市況分析コメントの精度を高める手法を提案する。

1 はじめに

近年、証券市場における個人投資家の比重が増大しており、個人投資家に対して投資判断の支援をおこなう技術の必要性が高まっている。そのため、人工知能分野の手法や技術を、金融市場における様々な場面に応用することが期待されており、例えば、膨大な金融情報を分析して投資判断の支援を行う技術が注目されている。その一例として、日本銀行が毎月発行している「金融経済月報」や経済新聞記事をテキストマイニングの技術を用いて解析し、経済市場を分析する研究などが盛んに行われている [1][5]。

本研究では、毎月のファンドの月次運用報告書に記載される市況分析コメントを自動生成する手法を提案する。市況分析コメントとは、例えば「8月の国内株式市況は、中国の景気減速懸念が台頭したことなどを背景とした世界的な株安を受けて大きく下落しました。」のような、その月における株価が大きく変動したイベント（例えば、「人民元の基準値切り下げ」）について述べ、株価が変動した理由を分析した文書である。以下に、2015年1月のファンド月次運用報告書に記載された市況分析コメントの一部を示す。

1月の国内株式市況は、ギリシャのユーロ離脱懸念やスイスフランの対ユーロ無制限介入終了による外国為替市場の混乱などを受けて大きく下落しましたが、ECB（欧州中央銀行）が量的緩和を導入するとの観測が強まると国内外の株式市況が反発、買い戻される展開となりました。...

上記のような市況分析コメントを記述するために、ファンド運用の担当者は①日経平均株価が大きく動いた記事を調べ、②その前後にあったイベントを確認し、③その記事の中から株価が変動した理由について述べている文を選択し、④まとめる、という作業を毎月、行う必要がある。これらの作業は、複数のファンドを担当する運用担当者にとって大きな負担となっているが、市況分析コメントについては運用担当者が独自の視点を生かして記述するというよりは、客観的に正確な記述をすることを求められるため、市況分析コメント作成については自動化を行い、自動化されたコメントの確認作業に専念した方が効率的な業務運営につながると思われる。

そこで、酒井らは、上記の①日経平均株価が大きく動いた記事の判定、②その前後にあったイベントの確認、③その記事の中から株価が変動した理由について述べている文を抽出、といった処理を自動化し、③の処理によって抽出された文を時系列順に並べることで市況分析コメントを自動生成する手法を提案した [8]。

*連絡先：成蹊大学 理工学部 情報科学科
〒180-8633 東京都武蔵野市吉祥寺北町 3-3-1
E-mail: h-sakai@st.seikei.ac.jp

†本論文に示された所見は著者自らのものであり、所属する組織の公的な立場を代表するものではない。

酒井らの手法の概要は以下のようになっている。

- Step 1:** 一か月分の日経新聞記事から日経平均株価について述べた記事（以降、分析記事と定義）を深層学習により抽出。
- Step 2:** 分析記事の集合から、その期間における重要なキーワード（以降、重要キーワード）を抽出（例：人民元、中国人民銀行）
- Step 3:** Step 1 で抽出した分析記事集合より、日経平均が変動した要因について述べた文（以降、要因文と定義）を抽出。
- Step 4:** Step 3 で抽出した要因文集合と、Step 2 で抽出した重要キーワード集合を使用し、重要な要因文を判定
- Step 5:** Step 4 で判定された重要な要因文を時系列順にならべ、市況分析コメントとする

しかし、酒井らの手法によって生成された市況分析コメントには、その月で起こった重要なイベントが含まれないことがあった。例えば、2015年1月において、「ギリシャ総選挙において反緊縮財政派が優勢」や「スイス国立銀行がスイスフランの対ユーロでの無制限介入終了を急遽発表」といった重要なイベントが自動生成したレポートに含まれておらず、Step 2においても重要なキーワードである「スイスフラン」「ギリシャ総選挙」なども取得されていなかった。

原因を調査したところ、入力とした2015年1月の分析記事において、スイスフラン高について述べている記事は1月16日の1つしかなく、重要なキーワードとして抽出されていなかった。同じく、「ギリシャ総選挙において反緊縮財政派が優勢」について述べている記事も1月26日の1つしかない。

本タスクは、複数の記事から1つの要約を生成する複数文書要約（例えば、文献 [6]）とみなすことができる。複数文書要約では、入力として複数の記事が与えられ、それらの記事から重要な文を抽出し、重要な文同士の冗長性を排除してまとめるという処理を行い、入力された記事の内容をまんべんなく含むような要約を生成することが求められる。複数文書要約では、通常、入力した複数の文書において多く、かつ、複数の文書にまんべんなく出現する語を重要キーワードと認定し、その重要キーワードから重要文を抽出するが、本タスクではこの仮定が成り立っていなかった。しかしながら、1月の（分析記事以外を含む）記事全体には、「ギリシャ総選挙において反緊縮財政派が優勢」に関する記事や「スイス国立銀行がスイスフランの対ユーロでの無制限介入終了」に関する記事は、それなりに多く出現していることが期待できる。そのため、本研究では、ある期間の分析記事のみから市況分析コメントを

生成するのではなく、分析記事に記載されているイベントを推定し、そのイベントに関連する記事を検索し、検索された関連記事をも使用して、市況分析コメントを生成する。

関連研究として、決算短信を対象として様々な情報を抽出する研究がある。酒井らは決算短信から業績要因を含む文（例えば「半導体製造装置の受注が好調でした。」）を抽出する手法を提案している [7][9]。坂地らは決算短信から原因・結果表現を抽出する手法を提案している [10]。北森らは決算短信から業績予測文（今後の業績予測に関する情報が記述されている文）を抽出する手法を提案している [2][3]。いずれの研究も抽出対象である情報を抽出するために有効な手がかり表現に着目し、それらの表現をブートストラップ的に自動的に獲得、もしくは、人手にて用意している。本研究においても、③その記事の中から株価が変動した理由について述べている文を抽出する処理のために手がかり表現を用いるが、その手がかり表現の獲得のために、酒井らの決算短信から業績要因を抽出する手法 [7] において提案された手がかり表現の自動獲得手法を用いている。

2 関連記事を用いた市況分析コメント自動生成手法

2.1 深層学習による分析記事の抽出

市況分析コメントを自動生成するための情報源として日経平均株価について言及している記事を使用する。そのため、一か月の日経新聞記事集合から日経平均株価について言及した記事（分析記事）を深層学習にて自動的に抽出する。分析記事の自動抽出手法については酒井らの手法 [8] と同一なので、手法の詳細は割愛するが、概要を述べると以下ようになる。

- Step 1** 2000年から2014年までの日本経済新聞記事のタイトルに「日経平均」が含まれている記事から、学習データを自動生成する
- Step 2** 自動生成された学習データから入力層の要素となる語（素性）を選択する。
- Step 3** Step 1 で生成した学習データと素性を用いて、深層学習にて一か月の日経新聞記事集合から分析記事を抽出する

2.2 関連記事の検索と重要キーワードの抽出

前節で抽出された分析記事に記述されているイベントについての記事（以降、関連記事と定義）を、その

月の記事集合から検索する。例えば、2015年1月の記事集合から、以下の記事が関連記事として検索される。

- 株・為替、波乱含みに、スイスフラン・ショックでユーロ安、欧州中銀、量的緩和が焦点
- 緩和マネー、日米欧株上昇、欧州量的緩和、市場に安心感、ギリシャ・原油なおリスク
- NY金、3週ぶり高値、ギリシャ政局混迷、買い先行

さらに、検索された関連記事集合から、その月における重要キーワードを抽出する。上記の手法により、2015年1月の関連記事から抽出した重要なキーワードの例を以下に示す。

原油安、ギリシャ総選挙、ECB、金属取引所相場、欧州中央銀行、原油先物相場、スイス国立銀行、スイスフラン

関連記事を用いることで、酒井らの手法では抽出できなかった「ギリシャ総選挙」や「スイスフラン」といった語が、重要キーワードとして抽出できるようになった。

2.3 分析記事からの要因文の抽出

分析記事から日経平均が大幅に変動した理由について言及している文（要因文）を抽出する。例えば、以下のような文を抽出する。

- ギリシャへの金融支援協議が難航していることや、中国・上海株の値動きへの警戒感から、投資家のリスクオフの動きが強まった。
- 米消費者物価指数の上昇やイエレン米連邦準備理事会議長が年内の利上げに前向きな姿勢を示し、為替相場が円安に向かった。

要因文の抽出は、「強まった」のような手がかり表現や、前節で抽出したその期間における重要なキーワードを使用して抽出する。手がかり表現の獲得手法は酒井らの手法 [7] と同一であるため割愛する¹。

分析記事からの要因文の抽出は、分析記事から手がかり表現と重要キーワードがともに含まれている文を抽出することで行い、さらに、重要キーワードに付与されたスコアを用いることで要因文にスコアを付与し、要因文のスコアが大きい順に抽出する。

¹ただし、最初に入力する初期手がかり表現は、本研究における抽出対象にあわせて変更している。

2.4 市況分析コメントの自動生成

前節で抽出された要因文を時系列順にならべ、市況分析コメントを生成する。しかし、抽出された要因文には同じ内容について記述されている文が含まれている可能性がある。そのため、抽出された要因文間の類似度を求め、類似度が大きい文の一方を除去することで、同一内容の文が市況分析コメントに含まれることを防止する。

例として、2015年1月の記事集合から生成した市況分析コメントを以下に示す。

今月の国内株式市況は日経平均株価が17408.71円から17674.39円まで上昇した。

前半は、17108.70円まで下落した。原油安やギリシャの政情不安などで5日の欧州や米国の株価が大幅に下落し、東京市場にも売りが波及した。業種別では原油安が響く鉱業の下げが目立った。資源国の経済への不安が高まったことで海外の金融・資本市場で動揺が広がり、東京市場にも及んでいる。ギリシャにユーロ離脱の懸念が浮上し、欧州経済の混乱への警戒感も株価を押し下げている。世界の市場を動揺させている原油価格の下落が止まらず、欧州経済懸念も強い。

後半は、17674.39円まで上昇した。原油価格の一段安やギリシャの政情不安が引き続き意識され、相場全体の先行き不透明感が増した。欧州中央銀行（ECB）の量的緩和観測を好感したものの、市場では止まらない金利低下への戸惑いも広がる。

第一文と「前半は...」「後半は...」から始まる文は、その月の日経平均株価の変動率から自動生成している。

2015年6月の記事集合から生成した市況分析コメントを以下に示す。後述の評価では、6月の市況分析コメントの評価値が最も高かった。

今月の国内株式市況は日経平均株価が 20569.87 円から 20235.73 円まで下落した。

前半は、20387.79 円まで下落した。米利上げ観測が再燃するのは直近の雇用統計の内容が良好で、景気回復期待が強まったからだ。

後半は、20235.73 円まで下落した。日本時間 18 日未明に米連邦公開市場委員会 (FOMC) の結果発表を控え、買いが手控えられた。FOMC ではメンバーの金利見通しが引き下げられ、市場で利上げが緩やかになるとの見方が強まった。前日の欧米株高やギリシャ債務問題の解決に向けた動きを好感した買いが入った。ギリシャの支援協議進展を好感したものだったが、再び交渉が難航すると相場は弱含んだ。今週は小売りを中心に 3~5 月期決算発表が本格化する。ギリシャの債務不履行が世界的な金融不安を招きかねないとの懸念が重荷になった。

2015 年 10 月の記事集合から生成した市況分析コメントを以下に示す。

今月の国内株式市況は日経平均株価が 17722.42 円から 19083.10 円まで急騰した。

前半は、18096.90 円まで上昇した。2 日発表の米雇用統計は低調な内容で、米国の早期利上げ観測がやや後退。中国など新興国経済の減速懸念から、9 月に入って企業業績への楽観論は急速に修正されてきた。前週末の米国株高に加え、環太平洋経済連携協定 (TPP) 交渉の進展が好感された。一部で追加緩和への思惑が浮上し、決定会合に向けて売り建てておいた先物を買戻す動きもみられた。9 月の中国製造業購買担当者景気指数 (PMI) の改善や米新車販売好調などでマクロ指標への不安が後退。10 月末には日銀が追加緩和するとの観測も強まる。

後半は、19083.10 円まで急騰した。中国人民銀行 (中央銀行) が 23 日に金融緩和策を発表し、中国景気への警戒感が後退。

3 評価

本手法の評価を行うため、本手法を実装した。実装にあたり、形態素解析器として MeCab²、係り受け解析器として CaboCha[4] を使用した。評価方法は、実際に運用担当者が作成した市況分析コメント (2015 年 1 月~12 月) と、その期間の日経新聞記事を使用して本手法により自動生成された市況分析コメントとを比較

²<http://taku910.github.io/mecab/>

表 1: 評価結果

期間	本手法	比較手法	運用担当者
1 月↑	0.47	0.42	0.72
2 月↑	0.41	0.39	0.39
3 月↑	0.54	0.42	0.74
4 月↑	0.40	0.38	0.60
5 月↑	0.56	0.38	0.40
6 月↓	0.67	0.45	0.60
7 月↑	0.56	0.45	0.47
8 月↓	0.51	0.41	0.88
9 月↓	0.46	0.50	0.88
10 月↑	0.49	0.33	0.87
11 月↑	0.43	0.57	0.69
12 月↓	0.48	0.63	0.46
平均	0.50	0.45	0.63

して行った。具体的には、ある期間における運用担当者が作成した市況分析コメントと自動生成された市況分析コメントとの類似度を、Doc2vec を用いて求める。そして、その期間における運用担当者が作成した市況分析コメントとの類似度の平均を評価値とした。ここで、比較手法として以下の手法と比較した。

比較手法: 酒井らの手法 [8] によって生成された市況分析コメント

運用担当者: 運用担当者が作成した市況分析コメントの 1 つを選択

運用担当者は理想的な結果に基づく評価値となる。結果を表 1 に示す。なお、期間の矢印 (↑, ↓) は、その期間において日経平均が上昇したか下落したかを示す。

4 考察

本手法と比較手法を比較すると、9 月、11 月、12 月を除く全ての月で、本手法がよい評価結果を得た。また、全期間の評価値の平均も本手法が上回り、その月のイベント内容に左右されずに、平均して高い評価値をだしていることが分かる。これは、本手法は、分析記事における記述が少ないイベントでも関連記事を使用することで、適切に重要キーワードへの重み付けができるからであると考えられる。また、比較手法では抽出された要因文同士の類似性の判定を行っておらず、類似した内容の文が生成された市況コメントに含まれている場合があったが、本手法では、類似している文の一方を除去することで、その月に起こったイベントを幅広く網羅している。しかし、そのことが評価値を下

げることもあった。例えば、比較手法のほうが高い評価値である12月の記事から生成された市況分析コメントを比較する。まずは、比較手法によって生成された市況分析コメントを示す。

前日の欧州中央銀行理事会で決まった追加の金融緩和策が市場の期待に届かず、失望売りが広がった。円高・ドル安が進んだのも重荷となり、輸出株の下落が目立った。幅広い範囲の業種に売り注文が広がり、東証1部上場企業の約9割が値下がりした。前週末の米国株の大幅下落に加え、15～16日には米連邦公開市場委員会が控え、日本株の先行きにも不透明感が高まっている。15日の東京株式市場で日経平均株価は大幅に続落し、ほぼ2カ月ぶりの安値を付けた。利上げの是非を判断する米連邦公開市場委員会の開催を前に投資家のリスク回避姿勢が強まり、東証1部の8割超の銘柄が下げた。

次に、本手法によって生成された市況分析コメントを示す。

今月の国内株式市況は日経平均株価が20012.40円から19033.71円まで下落した。前半は、18565.90円まで大幅に下落した。前日の欧州中央銀行（ECB）理事会で決まった追加の金融緩和策が市場の期待に届かず、失望売りが広がった。後半は、19033.71円まで上昇した。原油価格の下げ止まりも投資家心理を改善させた。16日の米連邦公開市場委員会（FOMC）の結果発表を控えて手じまい売りが続き、ほぼ全面安の展開。米連邦準備理事会（FRB）が利上げを決め、当面の金融政策の不透明感が払拭されたことが好感された。原油安が一服し、投資家が過度にリスクを避ける姿勢が和らいだ。2016年は低インフレ懸念が強まるとの見方が背景にあり、新興国の景気減速や通貨安が拍車をかけている。米利上げは米国の需要を抑えるだけでなく長短金利差が縮小し、銀行収益の圧迫要因となる。

この月は、「欧州中央銀行の追加金融緩和策」についてと、「FOMC（米連邦公開市場委員会）での政策金利引き上げ」についてが主な内容であり、それについてのみを記述した比較手法のほうが高い評価値を示した。本手法は、類似した要因文の一方を除去するため、基本的に1つのイベントは1つの要因文のみとなる。そのため、「原油価格の下げ止まり」といった多くのイベントが含まれた分析コメントとなり、結果として、評価値が低くなることもあると考える。

5 まとめ

本稿では、経済新聞記事などの経済テキストから、日経平均株価の市況について言及している記事のみを抽出し、それらの内容を自動的に要約することによりマーケットレポートにおける市況分析コメントを自動生成する手法について述べた。本手法では、まず経済新聞記事から深層学習により日経平均株価について言及している記事（分析記事）を抽出し、さらに、その分析記事と関連している記事（関連記事）を検索する。そして、関連記事を用いて重要なキーワードを抽出した。次に分析記事の中からその理由について言及している文（要因文）を抽出し、類似した要因文を除去しつつ、重要なキーワードを含む要因文を時系列順に並べることで市況分析コメントを自動生成した。評価の結果、関連記事を用いていない既存手法[8]より概ね高い評価値を達成した。

参考文献

- [1] 和泉潔, 後藤卓, 松井藤五郎: 経済テキスト情報を用いた長期的な市場動向推定, 情報処理学会論文誌, Vol. 52, No. 12, pp. 3309-3315 (2011).
- [2] 北森詩織, 酒井浩之, 坂地泰紀: 決算短信 PDF からの業績予測文の抽出, 電子情報通信学会論文誌 D, Vol. J100-D, No. 2, pp. 150-161 (2017).
- [3] Kitamori, S., Sakai, H. and Sakaji, H.: Extraction of sentences concerning business performance forecast and economic forecast from summaries of financial statements by deep learning, *IEEE Symposium on Computational Intelligence for Financial Engineering & Economics*, pp. 67-73 (2017).
- [4] 工藤拓, 松本裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol. 43, No. 6, pp. 1834-1842 (2002).
- [5] 藏本貴久, 和泉潔, 吉村忍, 石田智也, 中嶋啓浩, 松井藤五郎, 吉田稔, 中川裕志: 新聞記事のテキストマイニングによる長期市場動向の分析, 人工知能学会論文誌, Vol. 28, No. 3, pp. 291-296 (2013).
- [6] 酒井浩之, 増山繁: ユーザの要約要求を反映するためにユーザとのインタラクションを導入した複数文書要約システム, 日本知能情報ファジィ学会誌, Vol. 18, No. 2, pp. 265-279 (2006).
- [7] 酒井浩之, 西沢裕子, 松並祥吾, 坂地泰紀: 企業の決算短信 PDF からの業績要因の抽出, 人工知能学会論文誌, Vol. J98-D, No. 5, pp. 172-182 (2015).

- [8] 酒井浩之, 坂地泰紀, 和泉潔, 松井藤五郎, 入江圭太郎: 経済テキストからの市況分析コメントの自動生成, 第 20 回 金融情報学研究会, pp. 44–49 (2018).
- [9] 酒井浩之, 松下和暉, 北島良三: 学習データの自動生成による決算短信からの業績要因文の抽出, 日本知能情報ファジィ学会誌 (2019(採録決定)).
- [10] 坂地泰紀, 酒井浩之, 増山繁: 決算短信 PDF からの原因・結果表現の抽出, 電子情報通信学会論文誌 D, Vol. J98-D, No. 5, pp. 811–822 (2015).